DOCUMENT RESUME

ED 300 387 TM 012 202

AUTHOR Goldberg, Gail Lynn; Walker-Bartnick, Leslie
TITLE Maintaining Scoring Standards over a Rubric Transition Process.
PUB DATE Apr 88
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Descriptive (141) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Assessment; Elementary Secondary Education; Error of Measurement; Grading; *Scoring; *State Standards; *Testing Problems; True Scores; *Writing Evaluation; Writing Skills
IDENTIFIERS *Drift (Scoring); Maryland Functional Writing Test; *Rubric Transition Process; Score Variation; Standard Setting

ABSTRACT

A scoring rubric transition study is described. It was designed to evaluate possible drift in scoring the Maryland Writing Test from year to year (when using a modified holistic scoring method), to evaluate strategies for revising swing rubrics from narrative and explanatory writing while maintaining original scoring standards, and to establish strategies to control drift in the future. Drift involves a systematic change in the application of an observational scale or rating standard. A revised scoring rubric was constructed by committees of Maryland educators. These revised rubrics were then applied to rescoring papers from a 1987 equating study, using prompts from 1984. This return to responses based on the same prompt, recently scored on old rubrics, allowed the identification and correction of drift in scoring standards. Three pools of papers, each about 200 papers, were assembled to provide data on three hypotheses: (1) papers scored in 1984--the standard-setting year--would score lower in 1987 but would receive scores near the originals with the new rubric; (2) the revised rubric would give higher scores on the 1987 papers than would the old rubric; and (3) papers written in 1987 to 1987 prompts would receive lower scores when scored with the rubric before revision than with the revised rubric. These hypotheses were supported, indicating that drift had occurred and had been largely corrected with the revised rubric. It may not be possible to correct back fully, but scores and decisions are much more likely to be true from year to year. Five data tables provide the scoring results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED300387

MAINTAINING SCORING STANDARDS OVER A RUBRIC TRANSITION PROCESS

Gail Lynn Goldberg

Leslie Walker-Bartnick

Maryland State Department of Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G.A. Crenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, April 5-9, 1988, New Orleans, Louisiana

TM 012 202

Maintaining Scoring Standards Over a Rubric Transition Process

Gail Lynn Goldberg and Leslie Walker-Bartnick
Maryland State Department of Education

Introduction

In 1987, after a statewide writing assessment program was in operation for several years, MSDE undertook a multifaceted scoring study. The purposes of this study were three-fold: 1) To evaluate possible scoring drift from year to year when using a modified holistic scoring method; 2) to demonstrate and evaluate strategies for revising scoring rubrics for narrative and explanatory writing while maintaining original scoring standards; and 3) to establish strategies based on this information which could be used to control drift in future years.

Consentual observer drift is defined by Kazdin (1977) as "the tendency of observers to change the manner in which they apply the definitions of behavior over time." Since the phenomenon involves a systematic change in the application of an observational scale or rating standard, it will impact scoring accuracy without affecting interrarter reliability. Johnson and Bolstad (1973) reported that consentual observer drift is found in virtually all observational situations; therefore the goal of research must be to find ways to reduce or, optimally, eliminate it. Several authors have found ways to reduce or control consensual observer drift during scoring or observational periods (Taplin and Reid, 1973; Romanczyk et. al., 1973; Demaster et. al., 1977). Their findings led Kazdin (1977) to suggest raters should be trained together using criterion observations as examples, ratings should be made in random order and raters should be retrained periodically. The Maryland Writing Test uses all of these strategies for scoring a given administration of the test; however, no studies had been done to determine potential or actual drift (from year to year), although strategies such as equating studies were intended to ensure the stability of the score scale. There was some perception among teachers that the test had become more difficult, and there were several possible factors which might have caused a change in the application of the scoring standard. These include the continual development of the scoring guide over the first few years in which the test was given, the use of a different scoring contractor in 1985, and changes in student writing due to instruction. Recognizing these and other potential causes of drift, the rubric transition studies provided a means of determining if drift had occurred and

measuring the effects of strategies designed to reduce or eliminate drift.

The rubric transition scoring study was an outgrowth of various activities intended to refine the rubrics used to score the Maryland Writing Test. Introduced in 1984, the test requires that students respond to a pair of three-paragraph prompts, one narrative and one explanatory. Each response is scored on a domain-specific rubric covering five criteria--content, organization, audience, sentence formation and conventions of edited American English. In 1986, upon request by the Maryland State Department of Education, the Educational Testing Service (ETS) conducted a review of the MWT. While the program was identified as "state of the art," one suggestion made to improve the program was that the rubrics be refined to ensure greater clarity, consistency and validity of scoring. Because under any circumstances successful writing assessment programs must attend to the possibility of scoring drift from test administration to test administration, the ETS recommendation created a special challenge: to execute revisions in such a way that any drift from the original and standard setting year of 1984 be corrected and that the revised rubrics be true to that original standard.

Methodology: Rubric Revision

Over an eight month period, several groups of Maryland educators, including a Task Force on the MWT, the Test Specifications Committee and the Scoring Committee, contributed to the process of refining the scoring rubrics. Foremost among their considerations were the following:

--They wished to improve the match between the MWT modified holistic scoring criteria and the spirit and substance of writing instruction and evaluation in the classroom. There was a concern that a more specific link be forged between instruction and the way the five criteria of the scoring guide are interpreted or described at each point on the four-point scale.

--They wished to insure the clear and consistent applicability of the scoring criteria to the widest conceivable range of MWT prompts. While guide papers routinely change from test to test, scoring standards ought not be prompt-specific although clarifications of score decisions might conceivably vary depending upon the rhetorical demands of the prompt.

--They wished to develop greater parallelism across score points, so the degree of different features of writing, rather than differences in the features themselves, would clearly be the basis for attribution of a given score point. One consequence of this concern was the creation of

a MWT glossary so that students, parents, educators, and those scoring the writing test had access to identical information outlining criteria and the features of writing which contribute to each score point.

Revisions were tentatively made through a process which involved a wide constituency from the educational community. The final and most critical stage in the revision process required the reconvening of the Scoring Committee. This statewide group of twenty-four English teachers and Language Arts Specialists meets annually, before the operational test, to anchor to the original test and prescore responses to be used as training papers. SC members applied the revised rubrics and considered any additional changes or refinements needed. The central issue for the group was one of comparability. To ensure that no change in the language of the rubrics would result in a shift in score based on the original rubrics, SC examined the benchmark papers from 1984, and reannotated those papers using the language of the revised rubric. Where necessary, the form or language of the rubric was adjusted slightly to accommodate the original score decision.

Satisfied that the language of the revised rubrics yielded score decisions comparable to those of 1984, the committee's next task was to apply the revised rubrics to another pool of papers. For this purpose, papers were drawn from the 1987 equating study, which used the '84 prompts. There were four reasons for rescoring these papers with the revised rubric.

--It was deemed inadvisable to return to the benchmark papers of 1984 for retraining purposes because of significant changes in writing performance over the intervening years. Readers regarded the writers of 1984 responses as handicapped by the shorter booklet which permitted a maximum of a page and a half of text per response as compared to 3 pages since 1985. In addition, because of the change in the instructions in 1985 from "write a paragraph," to "write a paragraph or more," 1984 papers did not provide the best possible models.

--The return to responses which were based on the same prompt, which had been recently scored on the old rubrics, permitted the identification and correction of any drift in scoring standards which had been perceived but never statistically documented.

--The use of responses written in 1987 allowed comparison with 1987 operational papers, written to a different prompt. Any observable movement in scores as a consequence of the rubric revision could thus be regarded as extending beyond a particular prompt pair.

--Finally, the first application of the revised rubric by trained readers was to take place in the spring of 1987, when field tested prompts would be scored in preparation for the selection of an operational prompt pair for the 1988 MWT. Included among the new prompts during the field test were both the 1984 and 1987 prompt pairs. The application of the revised rubric by SC to pools of both 84 EQ papers and 87 OP papers provided a large sample of responses for use as training papers for field test readers. Since field test training materials were customarily drawn from papers prescored only by the MSDE specialists and scoring contractor's senior staff, these training papers had increased validity because they bore the clear agreement by SC using the revised rubric.

Methodology: Scoring

The hypotheses of this study were that: 1) papers originally written and scored during the standard setting year (1984) would receive lower scores when scored using the same rubric in 1987, but these scores would approach their original scores when rescored using the revised rubric 2) papers written in 1987 to the 1984 prompts would receive higher scores when the revised rubric was used than when these papers were scored in 1987 on the rubric before revision; 3) papers written in 1987 to the 1987 prompts would receive lower scores when scored with the rubric before revision than with the revised rubric.

Three pools of papers were assembled to provide data relative to the three hypotheses above.

Pool 1 This pool was assembled to test hypothesis number one. It consisted of approximately 200 operational 1984 papers which had been prescored by Scoring Committee and selected as training papers for scoring the 1984 MWT. As training papers, instead of reflecting a random distribution of score points this pool included a disproportionate number of borderline papers, high and low, at each line, to train readers to correctly discriminate between score points. In addition, because of their status as training papers, all of the original 1984 papers were identified by a solid "true score." Even "close calls" and line papers had to be identified by one of the four score points when first scored in 1984 by the Scoring Committee. The preponderance of line papers in this pool suggested that it was likely to be particularly sensitive to slight drift in scoring. Since MSDE's other procedures, including equating, indicated that MWT scoring was generally valid over the years, a less select pool, that is one with more solid scores, was likely to reveal little or no change in scores.

In this study, three scores were recorded for each paper in Pool 1: the original score assigned by SC in 1984, the

score assigned when the papers were rescored using the original rubric in 1987, and the score assigned when the papers were rescored using the revised rubric.

Pool 2 In order to test hypothesis two, a pool of approximately 200 papers written in 1987 to the 1984 prompts (as part of the equating study) was randomly selected. Two scoress were recorded for each of these papers: the original score assigned in 1987 using the original rubric, and the score assigned when the paper was rescored with the revised rubric.

Pool 3 The third pool of 200 papers were operational papers written in 1987 to the 1987 prompts. These papers were scored using the original rubric, and were then rescored using the revised rubric to provide the data to test hypothesis three.

All papers were scored during regular scoring, and it was not possible for readers to identify specific pools (in fact, readers did not know a study was being conducted). Scoring of all papers with the revised rubric was accomplished during spring, 1987 when field tested prompts were scored. All analyses were completed during fall, 1987. Subsequent studies involving rescoring these pools during 1988 operational scoring are in progress.

Results

Pool 1 Comparison of scores given to papers in pool 1 (papers written in 1984 to '84 prompts) in 1984 and 1987 reveals that in both domains, narrative and explanatory, mean scores dropped when papers were scored on the '87 rubric and increased somewhat (but not fully back to '84 means) when rescored on the revised rubrics (See Table 1). This provides evidence that drift from the 1984 standard had occurred, and that some papers were receiving lower scores in 1987 than in 1984. Close reexamination of these papers which received score changes revealed that they were for the most part those responses which illustrated the low or high end of a score point rather than a solid score. From 1984 to 1987, the scores for 45 papers in narrative and 62 papers in explanatory went down; most of the decrease was on papers which had been given scores of 3 or 4 in 1984. Upon rescoring with the revised rubric, the scores for 26 papers in narrative and 45 papers in explanatory went up. Over the three scorings, in narrative 38 out of 98 papers remained unchanged in score, 16 which had dropped in score increased back partly or fully to the 1984 score, and 26 which decreased when scored in 1987 remained at the lower score upon rescoring with the revised rubric. Of the 100 explanatroy papers, 41 exhibited the recovery trend, 28 remained unchanged, and 18 which were scored lower in 1987

remained at the 1987 score when rescored with the revised rubric.

Pool 2 Comparison of scores given to papers in pool 2 (papers written in 1987 to the 1984 prompts) shows that for both domains, the mean score based on the revised rubric were higher than scores assigned during 1987 scoring (see Table 2). In the narrative domain, 38 of the 99 papers were scored higher with the revised rubric. Six of the 38 (16%) had originally been scored a 2, 7 (18%) had originally been scored 2.5, 12 (32%) had originally been scored a 3, and 10 (26%) had originally been scored a 3.5. Of the 98 explanatory papers, 49 were scored higher with the revised rubric. Eighteen of the 49 (37%) had received a score of 2 in 1987, 16 (33%) had received a score of 2.5, 6 (10%) had received a score of 3.0 and 6 (10%) had received a score of 3.5.

The analysis of split scores revealed that 1987 scoring of the 99 narrative paprs resulted in 27 split scores, which decreased to 11 when the same papers were scored with the revised rubric. In the explanatory domain (98 papers) 1987 scoring resulted in a total of 33 split scores. Scoring with the revised rubric resulted in a decrease to only 7 split scores. It would appear from these data that the revised rubric does provide clearer, more consistent scoring while maintaining the original 1984 standard.

Pool 3 Analysis of score changes in pool 3 papers (written in 1987 to 1987 prompts) allowed us to determine whether the impact on scores of the revision to the rubrics would generalize to other prompts. Comparisions of scores assigned to these papers during 1987 scoring and again using the revised rubric reveal the same trends as the other two pools; the revised rubric scores were higher. (See Table 3) In the narrative domain, 66 of the 165 papers were scored higher with the revised rubric. Almost half of these 66 papers (32) were given a score of 3 in original 1987 scoring, another 14 (21%) had originally been scored a 2, and 7 (11%) had originally been scored as a 2.5. In explanatory, 40 of the 162 papers were scored higher with the revised rubric. Fourteen of these 40 papers (35%) originally received a 2.0, 10 (25%) originally received a 3.0 and 5 (13%) originally received scores of 1.0 and 2.5, each.

An analysis of split scores indicates that 1987 scoring of narrative papers resulted in 16 split scores; this number increased to 26 when papers were rescored with the revised rubric. 1987 Scoring of explanatory papers resulted in 36 split scores, which decreased slightly to 33 when papers were rescored with the revised rubric. Unique to this pool was the elimination of all the split scores which originally occurred during 1987 scoring. When split scores occurred

using the revised rubric, they were all new splits. These occurred because solid scores moved up by half a score point, (because one of two readers gave a higher score) reflecting the anticipated correction upward with the revised rubric. Upon reexamination, a number of these papers were seen as ones which, with future training on the revised rubric and with appropriate benchmark papers, would move to the higher solid score point.

Table 4 shows a summary of scoring for each pool of narrative papers by score point. The total percentage of papers with unchanged scores remained fairly stable over all three pools, as did the total percentage of papers with scores which went up and down. Many of the changed scores were originally scored as 1.5, 2.5 or 3.5, and the revised rubric was expected to provide clearer guidelines for scoring "close call" papers. It would appear that this modification of the rubric was successful. At the solid score points, scores were much more likely to remain unchanged. Finally, this table shows that when scores changed, they were more likely to go up when scored with the revised rubric. Since drift has been demonstrated to exist, this provides evidence that the revised narrative rubric begins to accomplish the goal of approaching the 1984 standard.

Table 5 shows the same summary results for the explanatory domain. Again, the 1.5, 2.5 and 3.5 score point results show that these are the original scores most likely to change as a result of the application of the revised rubric. The table also shows that scores are more likely to go up than down, evidence for correction of drift.

Discussion and Conclusion

A central aim in any assessment program is to ensure fair and stable scoring from administration to administration. Even with annual retraining before scoring and selection of anchor papers for specific test, and in spite of equating procedures, it is possible for slight drift to occur. The process of rescoring archival papers by various groups of readers provided a means of checking for drift and identifying rhetorical, formal and stylistic characteristics of student responses which contribute to a tendency for drift to occur. In addition, the transition studies yielded a sound pool of training materials which provide continuity between the original application of scoring standards and application of those standards using the revised rubrics. It is hoped that this study will contribute to the pursuit of greater score stability and validity by other programs in writing assessment.

As an outgrowth of the analyses (both statistical and textual) of score changes which resulted from the rubric

transition process, various steps have been implemented to facilitate complete correction back to the original scoring standard during scoring of the 1988 operational MWT.

--Rescored responses which did not return to their original '84 score were identified and inserted along with '84 EQ papers used to retrain Scoring Committee members as part of the anchoring process before they scored responses to the '88 prompts to use as training papers. In addition, the Writing Assessment Specialists and contractor's senior staff prescored field test responses to the prompt pair ultimately selected for operational use in 1988. These responses were followed through field test scoring, and responses receiving discrepant scores (when reader scores were compared to MSDE/MI expert reader scores) were reexamined. Following training on the revised rubric anchor set (drawn from responses scored by SC last April), committee members scored several training sets of field test responses to the new operational prompt. By comparing these papers, and their scores, with those from the '84 EQ pool used for training, SC members were able to correct back to the line established in 1984, and to orient themselves for the ensuing Scoring Committee activities necessary for the identification of training papers.

--The introduction of archival papers, established as part of the rubric transition process, will continue on a regular basis with all scoring sessions for the MWT, both operational and field-test, in order to keep close track of any tendency to drift from the original scores. It is anticipated that rescoring of the '84 OP responses during '88 operational scoring will demonstrate the full impact of the rubric transition activities, including the pre-test retraining of scoring committee after identification and discussion of responses failing to correct back to original score at the time of field test scoring last spring. While it may not be possible to correct back fully, we feel that after the examination of this year's score results we may be able to say with far greater confidence that scores and the decisions they impact upon are true, from year to year in which the Maryland Writing Test is administered.

LWB18000

Table 1

Means and Standard Deviations of 1984 Operational Papers (Pool 1)

Scoring Condition	Domain	
	Narrative	Explanatory
1984 operational score		
M	2.55	2.56
SD	1.03	1.04
1987 operational score		
M	2.16	2.04
SD	.72	.73
Revised rubric		
M	2.30	2.32
SD	.74	.84
Total N	98	100

Table 3

Means and Standard Deviations of 1987 Operational Papers (Pool 3)

Scoring Condition	Domain	
	Narrative	Explanatory
1987 operational score		
M	2.81	2.62
SD	.91	.97
Revised rubric		
M	3.09	2.74
SD	.92	.99
t	8.38*	3.45*
Total N	165	162

*Note. $p < .001$

Table 4

Score Point Breakdown of Change from '87 Scoring to Scoring with Revised Narrative Rubric

1987 Score Point	POOL 1 N=99			POOL 2 N=99			POOL 3 N=165		
	un-changed	up	down	un-changed	up	down	un-changed	up	down
1.0	9(56)	7(44)	0	3(75)	1(25)	0	7(54)	6(46)	0
1.5	1(14)	5(71)	1(14)	1(20)	2(40)	2(40)	0	1(100)	0
2.0	28(80)	4(11)	3(9)	20(71)	6(21)	2(7)	29(67)	14(33)	0
2.5	5(38)	7(54)	1(8)	1(11)	7(78)	1(11)	0	7(78)	2(22)
3.0	21(88)	3(13)	9	19(61)	12(39)	0	21(40)	32(60)	0
3.5	0	0	2(100)	1(8)	10(77)	2(15)	0	6(100)	0
4.0	1(100)	0	0	6(67)	0	3(33)	39(97)	0	1(3)
TOTAL	65(66)	26(27)	7(7)	51(52)	38(38)	10(10)	96(58)	66(40)	3(2)

NOTE. Numbers in parentheses are percentages of total at the specified score point.

Table 5

Score Point Breakdown of Change from '87 Scoring to Scoring with Revised Rubric Explanatory Rubric

1987 Score Point	POOL 1 N=100			POOL 2 N=98			POOL 3 N=162		
	un-changed	up	down	un-changed	up	down	un-changed	up	down
1.0	15(79)	4(21)	0	2(67)	1(33)	0	18(78)	5(22)	0
1.5	1(8)	10(77)	2(15)	0	2(20)	5(71)	0	2(66)	1(33)
2.0	20(57)	14(40)	1(3)	25(28)	18(42)	0	24(59)	14(34)	3(7)
2.5	0	12(92)	1(8)	3(16)	16(84)	0	2(18)	5(45)	4(36)
3.0	11(73)	3(20)	1(7)	10(63)	6(63)	0	29(70)	10(24)	3(7)
3.5	0	2(50)	2(50)	1(14)	6(86)	0	5(42)	4(33)	3(25)
4.0	0	0	1(100)	3(100)	0	0	26(87)	0	4(13)
TOTAL	47(47)	45(45)	8(8)	44(45)	49(50)	5(5)	104(64)	40(25)	18(11)

NOTE. Numbers in parentheses are percentages of total at the specified score point.

References

Demaster, B.; Reid, J. and Twentyman, C. (1977). The effects of different amounts of feedback on observer's reliability. Behavior Therapy, 8 : 317-329.

Johnson, S. and Bolstad, O. (1973). Methodological issues in naturalistic observations: Some problems and solutions for field research. In Hammerlynck, L. and Mash, E. (eds.) Behavior Change: Methodology, Concepts, Practice. New York: Research Press.

Kazdin, A. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. Journal of Applied Behavior Analysis, 10: 141-150.

Romanczyk, R.; Kent, R.; Diament, C. and O'Leary, D. (1973). Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 6: 175-184.

SPSSX, Inc. (1986). SPSSX User's Guide, 2nd Edition. Chicago: McGraw-Hill Book Company.

Taplin, P. and Reid, J. (1973). Effects of instructional set and experimenter influence on observer reliability. Child Development, 44: 547-554.

LWB18000

MAINTAINING SCORING STANDARDS OVER A RUBRIC TRANSITION PROCESS

Gail Lynn Goldberg

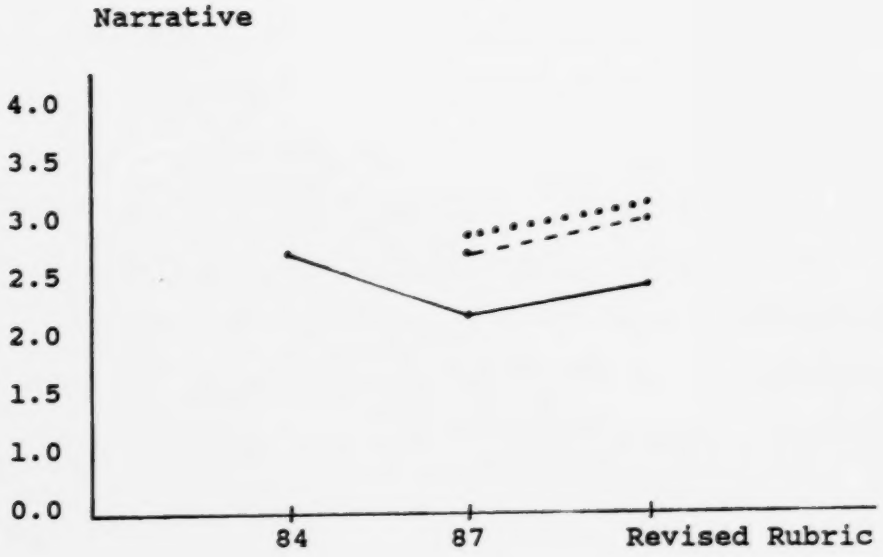
Leslie Walker-Bartnick

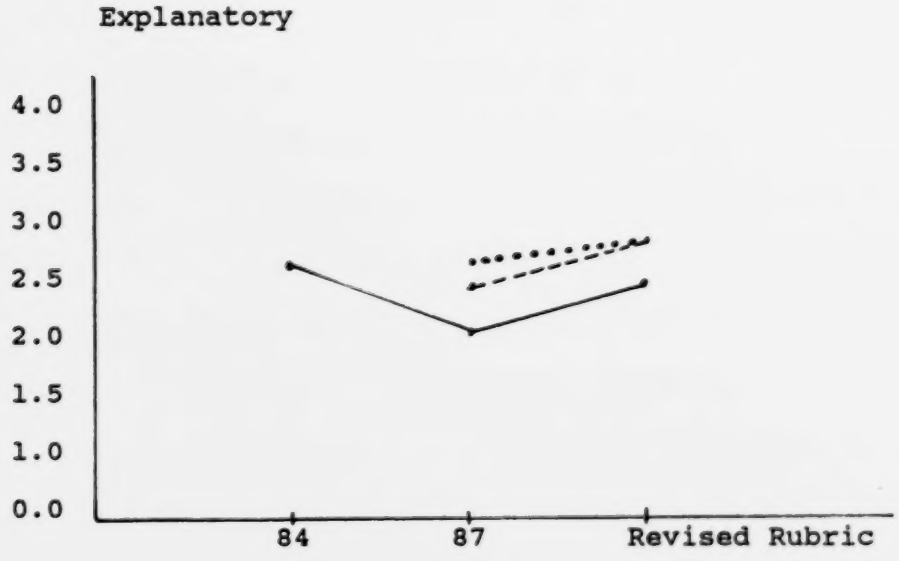
Maryland State Department of Education

Paper presented at the annual meeting of the American Educational Research Association, April 5-9, 1988, New Orleans, Louisiana

TM012202

Comparisons of Mean Scores, Pools 1, 2 and 3





——— POOL 1 (840P)
----- POOL 2 (87EQ)
••••••• POOL 3 (870P)

RESULTS - POOL 1

Analysis of Score Changes

Narrative (N=98)

	87 Result				Revised Rubric Result			
	84 Score	Unchanged	Up	Down	87 Score	Unchanged	Up	Down
	4	14	NA	20	4.0	1	NA	0
	3	10	0	20	3.5	0	0	2
	2	21	3	5	3.0	21	3	0
	1	1	4	NA	2.5	5	7	1
TOTAL		46	7	45	2.0	28	4	3
					1.5	1	5	1
					1.0	9	7	NA
					TOTAL	65	26	7

Trends - 84→87→Revised

38 papers
16 papers
26 papers

Explanatory (N=100)

	87 Result				Revised Rubric Result			
	84 Score	Unchanged	Up	Down	87 Score	Unchanged	Up	Down
	4	1	NA	21	4.0	0	NA	0
	3	2	1	28	3.5	0	2	2
	2	15	0	13	3.0	11	3	1
	1	16	3	NA	2.5	0	12	1
TOTAL		34	4	62	2.0	20	14	1
					1.5	1	10	2
					1.0	15	4	NA
					TOTAL	47	45	7

Trends - 84→87→Revised

28 papers
41 papers
18 papers

RESULTS - POOL 2

Analysis of Score Changes

Narrative (N=99)

Revised Rubric Result

87 Score	Unchanged	Up	Down
4.0	6	NA	3
3.5	1	10	2
3.0	19	12	0
2.5	1	7	1
2.0	20	6	2
1.5	1	2	2
1.0	3	1	NA
TOTAL	51	38	10

Explanatory (N=98)

Revised Rubric Result

87 Score	Unchanged	Up	Down
4.0	3	NA	0
3.5	1	6	0
3.0	10	6	0
2.5	3	16	0
2.0	25	18	0
1.5	0	2	5
1.0	2	1	NA
TOTAL	44	49	5

RESULTS - POOL 3

Analysis of Score Changes

Narrative (N=165)

Revised Rubric Result

87 Score	Unchanged	Up	Down
4.0	39	NA	1
3.5	0	6	0
3.0	21	32	0
2.5	0	7	2
2.0	29	14	0
1.5	0	1	0
1.0	7	6	NA
TOTAL	96	66	3

Explanatory (N=162)

Revised Rubric Result

87 Score	Unchanged	Up	Down
4.0	26	0	4
3.5	5	4	3
3.0	29	10	3
2.5	2	5	4
2.0	24	14	3
1.5	0	2	1
1.0	18	5	0
TOTAL	104	40	18